MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2019/2020

TDS2201 – STATISTICAL DATA ANALYSIS

(All sections / Groups)

17 OCTOBER 2019
9:00 am – 11:00 am
(2 Hours)

INSTRUCTION TO STUDENT

1. This question paper consists of 9 printed pages (inclusive of the front page) with 5 questions only. A formula sheet is attached (page 9). A booklet of statistical tables is provided separately.

2. Attempt **ALL 5** questions. Show **ALL** of your working steps clearly.

3. Please write your student ID in the space at the top right corner on this front page.

4. Please write all your answers in the space provided for each question in this question paper.

5. Your final answer for any calculation should be presented to **2** decimal values, unless it is value read from statistical tables.

QUESTION 1 [6 marks]

A researcher is interested to determine how much a student in MMU Cyberjaya campus spent on mobile data per month. A random sample of 500 students is selected from the admission record. The researcher has interviewed these students to record their mobile data fees.

a) What characteristic of the population is of interest to the researcher? You must write your answer as specific as possible. [1 mark]

b) What information are provided by a statistical distribution of variable? [2 marks]

c) What is the suitable plot to determine the distribution of the characterictic measurements in part (a)? [1 mark]

d) Identify the population of measurement that is of interest to the researcher. You must write your answer as specific as possible. [1 mark]

e) Identify the sample. You must write your answer as specific as possible. [1 mark]

Continued .....

QUESTION 2 [4 marks]

Study shows that El Niño, the periodic warming of the Pacific Ocean west of South America, affects the monsoon rains in India. Here are the monsoon rain falls (in millimeters) recorded in India for the 16 strong El Niño years between 1950 and 2000:

653	669	717	651
858	628	760	957
604	781	698	790
790	896	872	740

a) Find the mean, median, standard deviation and range of the above rainfall data.
[2 marks]

b) Is the distribution of the rainfall data symmetry or skew? Justify your answer.
[2 marks]

QUESTION 3 [10 marks]

a) Name the plot that can be used to determine the association between 2 continuous variables. [1 mark]

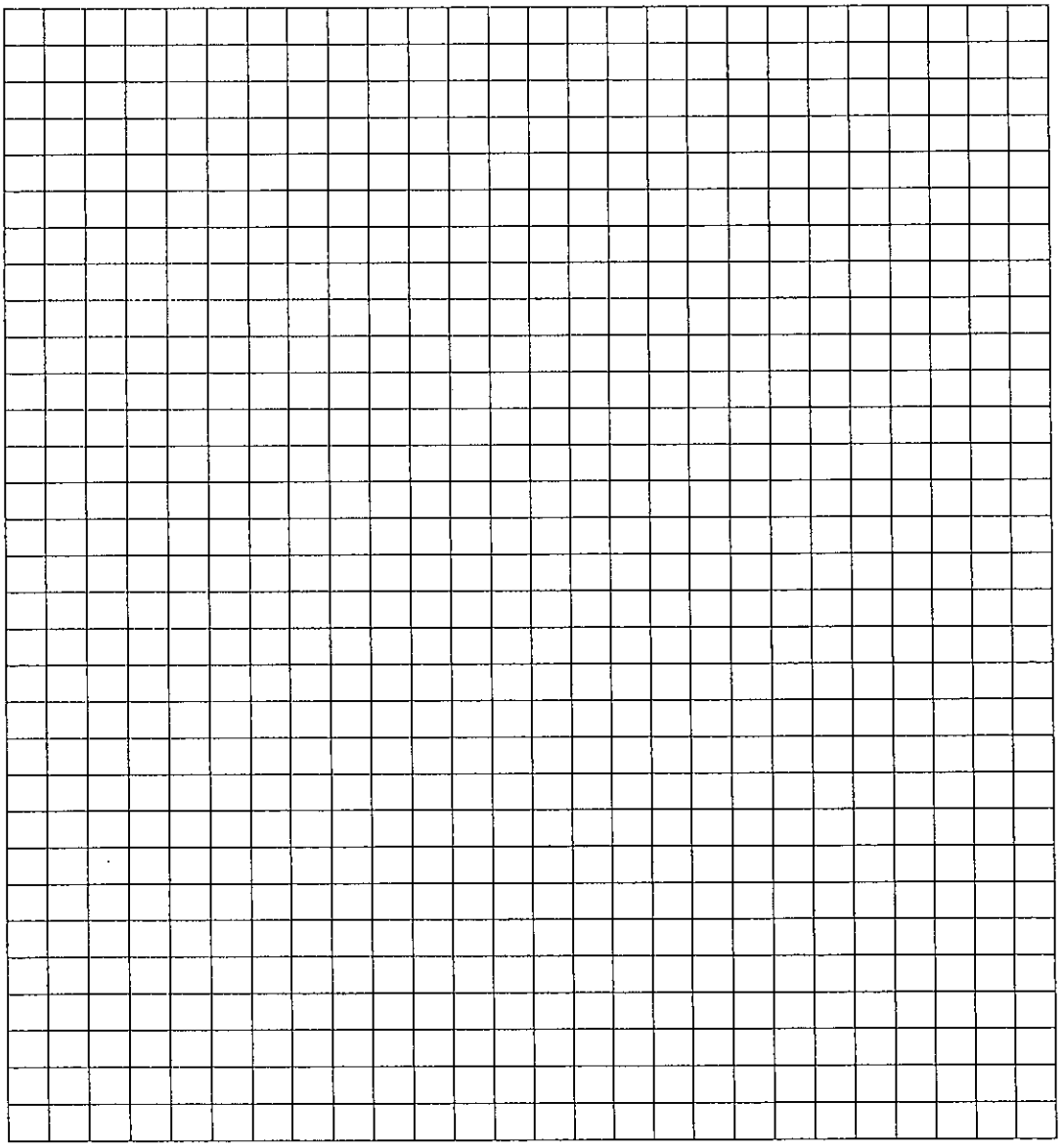
b) Name the plot that can be used to compare the distribution of a continuous variable for different categories of a categorical variable. [1 mark]

Continued .....

c) In the same study that collected the data in question 2, the monsoon rain falls (in millimeters) for the same 16 strong El Niño years between 1950 and 2000 are also recorded for Japan as below:

614	689	562	451
659	494	637	648
621	726	559	900
587	412	625	634

i) Construct a suitable plot to compare the distribution of monsoon rain falls in India and Japan. Your plot must be able to show the existence of outliers (if any) in the 2 data set. Ensure full labels of your axes. [4 marks]



Continued .....

ii) By referring to your plot in part (i), compare and contrast the distribution of the monsoon rain falls in India and Japan. [4 marks]

QUESTION 4 [10 marks]

The running times of films produced by two motion-picture companies are recorded as below:

Company	Running times (minutes)										
A	103	95	110	87	86	98	108	83	156		
B	85	99	83	142	97	88	118	161	133	109	122

a) Construct a 95% confidence interval for the variability ratio in the running times of the films produced by this two companies. Assume that the running times are normally distributed. [4 marks]

Continued .....

b) By referring to the 95% confidence interval in part (a), is there a difference in the variability of running times for films produced by company A and B? Justify your answer. [2 marks]

c) Referring to your answer in part (a) and (b), conduct an appropriate hypothesis test to verify if on average the film produced by company A has a shorter running time as compared to film produced by company B, with 0.05 level of significant. Write down clearly your null and alternative hypothesis and show all your steps. [4 marks]

Continued .....

QUESTION 5 [10 marks]

A study was conducted to determine the relationship between weekly advertising expenditures (dollars in thousands) and sales (dollars in thousands). The data is collected as follows:

Advertising expenditure	51	35	60	47	78	85	63	29	41	72
Sales	163	292	174	301	276	255	376	140	224	315

a) Advertising expenditures and sales are assumed to be linearly related. Fit a regression model for these two variables. [2 marks]

b) What happen to sales when a company spend an additional 1000 dollars in advertisement? [2 marks]

c) Estimate the sales when the average advertising expenditure is 4700. Calculate the residual of sales at this average advertising expenditure. You must write all your answers in unit dollar. [2 marks]

d) How much variability of sales is explained by the average advertising expenditure? Write down the name of the statistic you used to answer this question. [1 mark]

Continued .....

e) What is the assumption on residual when we fit a linear regression model? What plot can be used to check this assumption? Under what condition of the plot can we conclude that the assumption is violated? [3 marks]

End of Exam Questions.

Formula Sheet

Confidence Interval

1) $\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

2) $\bar{x} \pm t_{\alpha/2,\, n-1} \frac{s}{\sqrt{n}}$

3) $(\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

4) $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\, n_1+n_2-2} \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, $\quad s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$

5) $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, $\quad w_i = \frac{s_i^2}{n_i}$, $\quad \nu = \frac{(w_1 + w_2)^2}{\frac{w_1^2}{n_1 - 1} + \frac{w_2^2}{n_2 - 1}}$

5) $\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

6) $(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

7) $\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$

8) $\frac{s_1^2}{s_2^2} \cdot \frac{1}{f_{\alpha/2,\, \nu_1,\, \nu_2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot f_{\alpha/2,\, \nu_2,\, \nu_1}$

Test of significance

1) $\text{Test statistic} = \frac{\text{sample estimate - null value}}{\text{standard deviation of sample estimate}}$

2) $X^2 = \sum\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$